Peer Review and Access Models for Large-Scale Scientific Instruments

Kristen DeCaires Gall[†] and Diego A. Martin[‡]

June 10, 2022

Speeding Research on the Information Environment
researchaccelerator.org

Abstract

We examined the organizational structure, peer review, and access models for 17 large-scale scientific instruments from 13 different research organizations across multiple fields, including astronomy, the social sciences, climate science, nuclear physics, medicine, and machine learning.[1] Eight of the 17 instruments had a multinational leadership model. Five were housed at academic research centers and four within national governments. Eleven instruments had some sort of limitation on usage, usually incorporating an application process or affiliation by membership. We present two case studies that demonstrate thoughtful, inclusive, and seemingly effective processes for allocating usage of their respective instruments.

[1]The instrument access models examined are not exhaustive of all access models. However, the access models in this study provide *options* for the Institute for Research on the Information Environment to learn from and possibly imitate.

ł Author is affiliated with Princeton University
ǂ Author is affiliated with Purdue University

Funding: This research was generously supported by Microsoft, Craig Newmark Philanthropies, the John S. and James L. Knight Foundation, and the William and Flora Hewlett Foundation. This work was initiated as part of a joint project between the Partnership for Countering Influence Operations at the Carnegie Endowment for International Peace and the Empirical Studies of Conflict Project at Princeton University. **Author contributions:** KY collected the data, drafted the manuscript, and co-wrote the report. AW provided oversight over the research process and co-wrote the report. **Data and materials availability:** All data required to evaluate the conclusions are present in the paper or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Executive Summary

1. We gathered data on 17 large-scale scientific instruments from 13 different research organizations.
2. We gathered data for the following fields:
    2.1. Organizational mission
    2.2. Organizational and staffing structure
    2.3. Membership criteria for governing body
    2.4. Access to instrument
    2.5. External restrictions
    2.6. Decision-making process on technical/scientific matters
3. Eight of the 17 instruments had a multilateral, multinational leadership model. Five were products from academic research centers, and four were housed within national governments.
4. Eleven instruments surveyed had some sort of limitation on usage, usually incorporating an application process or affiliation by membership. Of this subgroup, eight had membership requirements with a national or citizenship component. All had some resource allotment reserved for nonmembers.
5. Of the 17 instruments, the six that were freely available were virtual resources. The remaining 11 had a time or resource allocation for researchers, with most of that subset requiring a peer-reviewed application process to obtain access. Of these 11, nine were physical instruments (telescopes, a laboratory, etc.) with time and/or resource allocation limitations.
6. All analyzed tools, except for the public access instruments, used field experts in their instrument application process–incorporating them into either their governing body, selection process, or external peer review committees.
7. We present two case studies that demonstrate membership and access models that appear to encourage collaboration, innovation, and effective instrument usage.
8. We outline two common best practices for a research organization, such as IRIE, to consider when developing an access policy for their created scientific instruments:
    8.1. Tailor access policies to their instruments by answering key questions around needs, risks and benefits for their created instruments and their associated products.
    8.2. Incentivize membership through data use and publication privileges.

Introduction

Resource allocation is a common issue when considering access to large-scale scientific instruments serving the broader research community. Some instruments have physical limitations that only allow only a certain number of users at a time, necessitating a policy on time usage. Others might require a certain level of training to use safely. An instrument's usage or end product can also expose sensitive information, necessitating a confidentiality policy or layered levels of access. For these reasons and more, the scoping team at the Institute for Research on the Information Environment (IRIE) examined a sample of existing access models for large-scale scientific instruments to garner themes and best practices for our future use.

After meeting with policymakers and researchers, we created a list of instruments commonly mentioned as complex and available to the public to include in our data set. We developed our criteria based on these initial conversations. Our team also conducted independent Google searches from April to June 2022 to expand the list. All data within this report are publicly available and require no subsequent interviews with organizational staff.

Table 1: Scientific Instruments Included

Organization	Instrument(s)	Type
Centre d'Accès Sécurisé aux Données (CASD)	Secure access data center (CASD)	National (French) but advised by an international council
Conseil Européen pour la Recherche Nucléaire (CERN, European Council for Nuclear Research)	Particle accelerator complexes	Multinational
	High-quality, low-cost medical ventilator design	
	The CERN technology World Wide Web	
	The CERN Grid	
European Southern Observatory (ESO)	The La Silla Observatory, containing several optical telescopes with mirror diameters of up to 3.6 meters	Multinational
Leibniz Institute for the Social Sciences in Mannheim	GESIS (German Social Science Infrastructure Services) Panel	National (German)
Mount Stromlo and Siding Spring Observatories	A 2.3-meter telescope	Academic
	The Anglo-Australian Telescope (AAT)	

National Aeronautics and Space Administration (NASA)	The International Space Station	National (United States)
National Opinion Research Center (NORC) at the University of Chicago	The General Social Survey (GSS)	Academic
National Weather Service	National Digital Forecast Database Graphical Forecasts	National (United States)
North Carolina State University	Laboratory for Analytic Sciences	Academic
Sloan Digital Sky Survey	The Sloan Digital Sky Survey	Multinational
The Gemini Observatory	Twin 8.1-meter diameter optical/infrared telescopes	Multinational
National Human Genome Research Institute	The Human Genome Project	Multinational
University of California, Irvine	UCI Machine Learning Archive	Academic

Methodology

We gathered data on the 17 shared scientific instruments above for the following characteristics:

1. Organizational mission
2. Organizational and staffing structure
3. Membership criteria for governing body
4. Access to instrument
5. External restrictions
6. Decision-making process on technical/scientific matters

Our inclusion criteria were as follows:

1. Measurement of the phenomena of interest for the organization could not be done by individual researchers.
2. Data collected by the instrument were processed and made available to researchers who did not participate in building them.
3. The managing organization maintained the physical and computational infrastructure of these instruments, which independent researchers could then use.

Our list included physical instruments (such as telescopes) and intellectual ones (such as datasets). Instruments within our data supplied various functions to the research community: astronomy, social sciences, climate sciences, nuclear physics, medicine, and machine learning. Eight of the 17 instruments had a multilateral, multinational leadership model; five instruments were products from academic research centers, and four were from national governments (see Table 1).[2]

Governance

Six of the eight instruments with publicly available governance models had a council, board, or committee that oversees administrator(s) carrying out daily operations and decisions. The administrators were responsible for the organization, operations, and budget. In contrast, one of the eight instruments noted that members elected the board. The remaining instrument was governed and administered via a member committee.

Organizational Structure

Out of the 17 instruments studied, all had an executive or president, 15 had an advisory board or council, and 14 had subordinate directors to the president. Eight of the 17 instruments had member voting as an integral part of their organizational or operational structure. One example of the majority's approach is the GESIS Panel of the Leibniz Institute for the Social Sciences. They had a Scientific Advisory Board, Board of Trustees, president, subordinate directors for issues such as Survey Data Curation and Survey Design and Methodology, and a user advisory board.

Access

We gathered information on eligibility restrictions for instrument usage as well as the process to obtain it, if applicable.

Table 2: Data Access Categories across Instruments

Access Type	Instruments
Open	CERN's Technology World Wide Web The Human Genome Project The General Social Survey (GSS) National Digital Forecast Database Graphical Forecasts GESIS Panel UCI Machine Learning Archive
Licensure Required	CERN High-quality, low-cost medical ventilator design CERN Particle Accelerator Complexes

[2] We found seven instruments with some information but not enough details to include in the review.

Request for Proposal/Application Process / Membership	Mount Stromlo: 2.3-meter telescope Gemini Observatory: 8.1-meter diameter optical/infrared telescopes GESIS Panel (extended version) Laboratory for Analytic Sciences The La Silla Observatory Telescopes CASD Secure Data Access Center and Microdata CERN's Worldwide LHC Computer Grid The Anglo-Australian Telescope (AAT) Sloan Digital Sky Survey International Space Station

Most instruments examined had some sort of limitation on usage, usually incorporating an application process or affiliation by membership. Of this subgroup, eight had membership requirements with a national or citizenship component, meaning the organization had a clear national or governmental affiliation or existed as part of a predetermined alliance of countries. Within this subset, all had some kind of loophole or resource allotment reserved for nonmembers. For instance,

1. Mount Stromlo and Siding Spring Observatories allowed open access nights to its 2.3-meter telescope to all Australian astronomers who completed a successful usage application. Non-Australian teams could still apply but had to pay for usage. A committee met quarterly to discuss applications and time assignments for using the telescope. Access to its AAT telescope was limited to astronomers who were affiliated with member institutions.
2. The Gemini Observatory had a bi-annual call and review for usage proposals. Any astronomer from one of its five partner countries (Argentina, Brazil, Canada, United States, and South Korea) could apply to use its telescope. However, 10% of the telescope time was reserved for researchers outside of the member countries.
3. CASD consisted of six Research Data Centers from Germany, France, the Netherlands, and the UK. All access to its data required a peer-reviewed application. CASD explicitly stated its commitment to facilitating open access to data for researchers across the EU, and more recently North America, with certain provisions.
4. The North Carolina State University Laboratory for Analytic Sciences had an annual call for white papers detailing proposals for immersive collaboration with their facilities and a 12-month funded effort. Applications were reviewed by the lab's principal investigator team.

Peer Review Process for Instrument Access Requests

Out of the 17 instruments, 6 were freely available, and the remaining 11 had a process for access or project approval. Those 11 included an application as part of the request for a project proposal. Five included an explicit peer review process, including the La Silla Observatory, which reviewed requests through a dual anonymous peer review.

All tools examined, except for the freely available public instruments, used field experts in their peer review process for instrument usage, either by way of incorporating them into their governing body or selection process or through the use of external peer review committees. Expert panels could include a scientific council, a group of principal investigators (mainly in the case of academic models), or members of a Knowledge Transfer team, as in the case of CERN licensing applications. While most of the organizations defined the review process for accessing their instrument, the exact makeup and/or membership of the deciding body was not always available. Additional interviews with organizational staff could provide more insight into process details.

We categorized three of the fourteen processes as requiring organizational training for access. For instance, to access the GESIS Panel, applicants had to review provided literature on using the data prior to approval. For NC State University's Laboratory for Analytic Science, the RFP process implied that support and access would be part of a larger partnership, and that NC State would be heavily involved in the project's oversight.

Four instruments required resources other than a PC to access. For example, access to most CERN Grid data (other than Tier 3) required high-performance computing. CASD sent approved applicants data on an SD-Box, an autonomous terminal allowing secure access to its central IT infrastructure.. Three organizations included output review, and four required that approved applicants sign a contract, such as a data use agreement. Two organizations, NASA and NC State University, also included requests for funding as part of their application process.

Below, we highlight two instruments with complex project teams, inclusive governance, and peer-review frameworks.

Case Study 1: CERN

CERN is one of Europe's oldest joint research ventures, established in 1951 by a multinational meeting of UNESCO. Its most notable physical instrument is the world's largest and most powerful particle accelerator, along with its accompanying data (a.k.a. "the CERN Grid"). CERN has a demonstrated history of supplying the international community with cutting-edge technology–including the world's first website and subsequent publicly available World Wide Web in 1993, and more recently, a high-quality, low-cost medical ventilator model for use during the COVID-19 pandemic. These three instruments, in addition to others at CERN, are listed in our dataset.

CERN now has 23 member states in its Council, its highest authority. Each member contributes financially to CERN programming, gets a single vote on council decisions, and appoints two delegates, one representing scientific interests and the other national interests. All member states besides Israel and Greenland are part of the EU. Japan and the United States hold Observer status; Russia's was revoked in March 2022. The Council is advised by a Scientific Policy Committee, which can contain members outside of the member states, and a Finance Committee that oversees financial contributions. Members of the Scientific Policy Committee, in addition to being tasked with general advisement to the Council on all CERN scientific matters, review decisions related to particle accelerator and ventilator use. They are elected experts appointed by the Council based on scientific merit without reference to nationality and include those from non-member states.

CERN used three of this report's named access models for its different instruments, ranging from the freely available model for its World Wide Web to more restrictive models for its more sensitive technologies. CERN demonstrates how a multilateral organization can manage various levels of peer review under the same research umbrella.

Licensing Requirements for Replicable Technology

During the start of the COVID-19 pandemic, CERN temporarily shifted its focus to producing a replicable design for a high-quality, low-cost ventilator for international dissemination. CERN required licenses to access and exploit the prototype technology. Though the spirit of this endeavor was to make the design freely available, potential users had to submit proof that they were able to manufacture and distribute the ventilators effectively and include a development and commercialization plan for use in DAC-list countries (nations and territories eligible to receive developmental assistance by the Organization for Economic Co-operation and Development). While CERN did offer limited consultancy in distributing the technology for the ventilators, their language suggested that application outcomes would be most favorable for those who could take on all stages of production independently, including a regional assessment of the area to be served.

Separately, to use technologies associated with the particle accelerator, CERN encouraged research and development collaborations with its Knowledge Transfer group and the support of spin-off companies with CERN technology and consultancy. CERN owned all intellectual property resulting from any subsequent patent. Their website recommended reaching out to the Knowledge Transfer group as a first step.

Using Membership as an Access Point for Research Instruments

The CERN Worldwide LHC Computer Grid is a global collaboration of over 12,000 scientists spanning 40 countries and 170 computing centers, using 900,000 computer cores. Its network used a four-tier distribution system (0-3), with Tier 0 being CERN's Data Center housing mostly raw data and only consisting of about 20% of the grid's total capacity. Tier 1 consisted of 13 large computing centers with enough capacity to store LHC Data and provide round-the-clock servicing. Tier 2 was a group of 155 approved universities and scientific institutes with sufficient capacity to process Tier 1 data and share the production and reconstruction of simulated events.

Individual scientists could access Tier 3-level data from Tier 2 institutions. To gain access to CERN Grid information, one would need to be part of an assigned Virtual Organization (VO) designated by a CERN project. Each project's access level would be determined by the Scientific Policy Committee; each VO had access to only the information deemed necessary its for designated projects. An individual could agree to volunteer their personal computers to be part of Tier 3. There was an approval process to establish a new Virtual Organization and a registration process to become part of any Virtual Organization.

Case Study 2: Sloan Digital Sky Survey

The Sloan Digital Sky Survey is the most extensive map of the universe ever created to date. The project began regular operations in 2000 and, at the time of study, was the collaborative effort of dozens of research and educational institutions classified as Full Members, Associate Members, or Participant Groups. An Advisory Council governed the project and made recommendations to the overarching Astrophysical Research Consortium (ARC), which owned and managed the Apache Point Observatory, where the survey is conducted. All data released from the Sloan Digital Sky Survey (SDSS) was considered public domain.

The SDSS is notable for its extensive member list and publication and data-sharing policies. The SDSS Policies and Procedures required that every project undertaken using the SDSS must be posted online, "specifying the subject matter, project leader, known collaborators, a contact person, and the anticipated duration of the project. Lists of proposed projects and publications will be maintained by the SDSS-IV [current version] spokesperson and made available on the internal project website." Under this transparent design, SDSS members could make requests to work as an author on a given effort, assuming they have contributed significantly to it. Data from the project was only available to participants of member institutions before release in the public domain.

To access the SDSS data before the public, two models existed–one for SDSS members and one for others. Membership was usually granted to faculty or research-track scientists at full or associate member institutions, and data rights were extended to those members' students and postdoctoral researchers. Member researchers had to inform SDSS administration if a project was anticipated to lead to a publication and undergo output review. Members did not need to fundraise, as the member institutions or individuals have already "bought into" early access. In contrast, if not a member, the researcher had to:

1. Request comment and input from working groups (while not a requirement, this step is highly recommended);
2. Request External Collaborator status via form;
3. Gain approval; and
4. Undergo output review.

Best Practices

Tailoring Access to the Instrument

CERN used multiple access models for its instruments under the same umbrella organization based on the ability of the end user to responsibly and effectively consume its product. In the case of its high-quality, low cost ventilator used during the COVID-19 pandemic, CERN incorporated expertise in technology transfer, community needs assessment, and licensure to successfully deploy a highly sought-after tool to communities in need. For the CERN grid data, that same organization deployed a complicated, high-throughput computing system platform that had multiple access levels, required heavy infrastructure, and would not necessarily be usable by an individual researcher. CERN was able to effectively tailor access plans based on the product, rather than having a blanket policy for the entire organization.

Based on the CERN access models and the access models of the other 16 analyzed instruments, basic questions to consider when building a model of access for a research instrument and its products:

1. *Who needs to access the instrument?*
2. *Who needs access to its product(s)?*
3. *Are there benefits to putting limitations on access to the instrument but making its product(s) openly available?*
4. *Is expertise required to operate it?*
5. *Is there proprietary information involved, or are there any sort of confidentiality issues that should be considered?*
6. *Are there risks associated with using the instrument or consuming its product(s)?*
7. *Are there costs and/or benefits associated with a certain population gaining access to the instrument or its product before the general population?*

Answering these fundamental questions can help shape policies and infrastructure around scientific instrument usage for a new research organization.

For IRIE, we propose the following answers to the above questions:[3]

1. *Who needs to access the instrument?* Researchers from across the world from civil society, academia, and government. IRIE may choose various eligibility requirements, including an affiliation requirement.[4]
2. *Who needs access to its product(s)?* Other researchers, the public, the government, and social media platforms.
3. *Are there benefits to putting limitations on access to the instrument but making its product(s) openly available?* Yes. Depending on the instrument and its output, expertise may be required to use it safely and effectively. However, its product (i.e., data) can be responsibly distributed to save time, resources, and energy for communities that need it.
4. *Is expertise required to operate it?* Yes.
5. *Is there proprietary information involved, or are there any sort of confidentiality issues that should be considered?* Yes.
6. *Are there risks associated with using the instrument or consuming its product(s)?* Yes.
7. *Are there costs and/or benefits associated with a certain population gaining access to the instrument or its product before the general population?* Yes to both.

---

[3] The answers proposed for these questions are not definitive, but represent potential solutions based on IRIE's mission and vision for its future instrument.

[4] See brain trust notes; Reynolds, Jen Rosiere , Aditi Bawa, and Kamya Yadav. "Researcher Access to Restricted Government Data". *Carnegie Endowment for International Peace* (2 June 2022). https://drive.google.com/file/d/10dK79PbSWG5hrrgCvcsXfWeaiKxY7vZP/view?usp=sharing

Incentivizing and Developing Membership through Data Use and Publication Privileges

The Sloan Digital Sky Survey demonstrated a model where data and publication opportunities are available to members before data is available to the general public. For a new organization, this could be a way to entice new members and acquire buy-in, especially during the startup phase. Member institutions that bought into SDSS got early access to data collected in the survey and could use this data to write scientific papers. Membership type–either full or associate–determined how many participants from the institution can access the proprietary data.[5]

SDSS' Publication Policy governed publications based on data before their release to the general public and outlined an inclusive authorship policy. It stated, "A scientific paper will include in its author list any individual who has made a significant contribution to that specific research analysis, along with all Architects who have requested authorship.[6] Authorship requests should be made via the webform available through a link on the Publication Archive page. With the exception of Architects, those requesting co-authorship must include a brief description of their contributions to the project. The first author, in consultation with the analysis team, may determine an acknowledgment is more appropriate in some cases."[7] We did not find this model in any of the other instruments and associated research societies. However, IRIE could consider implementing a similar policy when working with in-house generated products from its instruments.

Conclusion

Most shared scientific instruments we studied had some sort of limitation on their usage. These access levels largely depended on the instrument and governing research organizations themselves and did not necessarily map evenly across similar instruments. But examining these practices is a useful exercise, as it helps us to share best practices for effective and safe resource allocation for shared scientific instruments. Organizations could tailor access policies to their own created instruments by answering key questions around needs, risks, and benefits for those instruments and associated products. Furthermore, researchers could use instrument access as a tool to enhance the membership experience for their organization.

---

[5] "Principles of Operation for SDSS-V". *Astrophysical Research Consortium* (12 December 2017). https://www.sdss5.org/wp-content/uploads/2022/04/SDSS-V-Principles-of-Operation.pdf

[6] "Architects are those who have made substantial contributions to the success of the SDSS-V survey outside of explicit scientific results, including (but not limited to) work on optics, telescope, infrastructure, calibration, camera, spectrographs, data reduction and archiving software, commissioning, management, collaboration climate, centrally organized or sponsored work in the areas of education and public outreach, and major fundraising."

[7] "Publication Policy". *SSDS-V* (2022). https://www.sdss5.org/collaboration/publication-policy/

Appendix

A.1 Codebook

Variable	Description
id	a number from 1 to 13 depending on the order we found the instruments.
type	type of the instrument (categories: national, multinational, academic, National (country), but advised by an international council)
group instrument	category of the instrument (e.g., databases, particle accelerator, medical ventilator, nuclear research, telescopes, space missintions, applications)
instrument	instrument name
name of the organization	name of the organization in charge of the instrument
mission	organizational mission presenting the business, products or services, and customers, as well as, the primary objective of the institution in charge of the instrument
organizational structure	overall structure of the organization
staffing structure	organization chart including the main job positions (e.g., director general) and the areas inside the institution (e.g., academic and professional staff)
membership criteria for governing body	how members are decided and the members that are part of the organization
access	how the instrument is available, prioritization management, and international patterns
process to access	steps to access the data including calls for using the instruments and general rules to work with the tools
external restrictions	some instruments have external restrictions such as license committee or councils, other instruments are freely available
decision-making	who made the decision of how to use the instrument
sources	each of the sources to build the data

A.2 Data

Access Models of Large-Scale Shared Scientific Instrument